DOCUMENT RESUME

ED 325 118 IR 053 242

AUTHOR Francis, Hannah; Liddy, Elizabeth D.
TITLE Structured Representation of Theoretical Abstracts: Implications for User Interface Design.
PUB DATE May 90
NOTE 13p.; Paper presented at the Mid-Year Meeting of the American Society for Information Science (San Diego, CA, May 21-24, 1989).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Abstracts; Computer System Design; Discourse Analysis; *Discourse Modes; *Information Retrieval; *Linguistic Theory; Man Machine Systems; Political Science; *Rhetorical Theory; Search Strategies; User Needs (Information)

ABSTRACT

A descriptive study was conducted to delineate a structured representation of one abstract type (theoretical) and suggest how this predictable representation could affect user-system interfaces of information retrieval systems. Fifty-five argumentative type abstracts were selected from issues of "International Political Science Abstracts." Each abstract was classified into one of three groupings: Substantive, Authoritative, or Motivational. Each abstract was examined for the presence of the first and the second triads of elements--i.e. (1) data, warrant, and claim; and (2) backing, rebuttal, and qualifier. Of the abstracts, 68% belonged to the Substantive class, 20% to the Authoritative class, and 11% to the Motivational class. The first triad of elements was found in all of the abstracts. Of the second triad of elements, Rebuttal was found in 16, Qualifier in 24, and Backing in 28 of the abstracts. It is concluded that patterns of form and structure can be found in the argumentation texts, and therefore, that the rhetorical structure of argument is a useful framework in which to describe the structure of theoretical abstracts. Further, if a user-system interface could make use of these findings in information retrieval systems, users would be aided in comprehending the content of retrieved abstracts and in refining their search queries. (8 references) (SD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED325118

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

STRUCTURED REPRESENTATION OF THEORETICAL ABSTRACTS: IMPLICATIONS FOR USER INTERFACE DESIGN

by

Hannah Francis

and

Elizabeth D. Liddy

May 1990

IR053242

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Elizabeth Liddy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Structured Representation of Theoretical Abstracts: Implications for User Interface Design[1]

Hannah Francis and Elizabeth D. Liddy
School of Information Studies
Syracuse University

INTRODUCTION

Information retrieval is concerned with finding those few documents in a system which best fit the users' current information need. As such, the retrieval process is concerned with two types of representations: statement of the users' information needs, problems or queries, and the documents or their surrogates stored in the retrieval system. Performance of the system could be upgraded by improving either of these representations.

There is much in the literature on users' problem statements (Taylor, 1968; MacMullin & Taylor, 1984; Belkin, Oddy & Brooks, 1982), to suggest that expecting more of users' statements is not realistic. As Taylor pointed out in his seminal paper (1968), having users describe what it is they need, when they themselves are hardly articulate about what it is they don't yet know, is putting quite a strain on the communication aspect of the retrieval process.

At the same time, document representation techniques have advanced very little since the last major innovation of making the natural language portion of documents, usually abstracts, available for free text searching. This is not to say that more could not be done to improve the document representations, but until fairly recently, the reliance within information retrieval on statistical techniques of content representation showed few opportunities for substantive improvement. In current retrieval systems, a variety of term frequency measures are used in an attempt to determine which documents might be more 'about' a user's topic of interest. Use of linguistic techniques, intuitively the most appropriate approach, has been limited to lower levels of linguistic analysis, namely morphology (stemming) and syntax (noun phrase identification). Recently, the feasibility of applying higher levels of linguistic processing, such as discourse analysis, for the improvement of document representation has been recognized. It may be possible for this new document representation to be of assistance to the user in increasing the understanding and statement of the information need.

In this paper, we report on efforts to delineate a structured representation of one abstract type and suggest how this improved representation may affect the user-system interface of retrieval systems. The impetus for this investigation was the highly

[1] The investigation reported herein was supported by a grant from the Council on Library Resources.

IR053242

encouraging results found by Liddy (1988) in describing the discourse-level structure of empirical abstracts. In the hope of extending this general approach to a quite distinct type of abstract, we sought to discover whether abstracts reporting on theoretical work have a predictable set of information components and whether these occur in any predictable ordering within abstracts.

DISCOURSE ANALYSIS

Since documents and abstracts are texts, they are amenable to discourse analysis, the level of linguistics which is concerned with how units of language larger than a sentence function. Discourse linguistics is concerned with how these texts communicate, both through their structural organization and through the meaning imparted by sentences interpreted as a text, rather than singly. One major aspect of discourse linguistics is the investigation of the nature of the implicit structural organization of information in texts of different types. This is referred to as either text-level or discourse-level structure. It can be thought of as a superstructural organization of semantic content. However, it is not devoid of meaning, for the structure itself implicitly communicates semantically. For example, the simple fact that a sentence is the last line of a story communicates that what it contains is to be interpreted as the end of the story.

The theory of discourse linguistics suggests that texts which serve a common purpose among a community of users eventually take on a rather predictable structure and organization. Readers of office memos, city ordinances, academic course descriptions, and even obituaries could all attest to this fact.

STRUCTURE AND FORM

The overriding concern here is a search for the structure and form of the rhetorical reasoning process that can assist in the understanding of the nature of argumentation as it appears in theoretical abstracts found in information retrieval systems. Conventional argumentation theorists have been influenced by, and have rooted their work in classical Aristotelean rhetorical proof. (Anderson & Dovre, 1968, p. 235). The essence of this form differentiates between the treatment of probabilities (enthemene) and that of certainties (syllogism).

A modern incarnation of this classical form is found in the work of McBurney and Mills (1968) who consider argumentation to be primarily concerned with levels of certainty. This orientation led the authors to develop a structure for analysis based on the inductive-deductive frame of reference. According to this model, argument is seen as a range of probabilities i.e. 'possible, probable, plausible or certainly true'. (McBurney & Mills, 1968 p. 240) The authors also make the case for the inductive process as a means of getting at deductive argument. Here, induction is used as a linkage in the reasoning sequence to get at deductive

arguments. Utilizing this philosophical base as a point of departure, McBurney and Mills developed a classification of arguments based on the logical relationship between the premise and conclusion. The forms of argument they identify are: cause, sign, example and analogy.

Yet another approach to argument understanding has been posited by Brockriede and Ehninger (1968). Employing the conceptual principles developed by Toulmin (1964), these authors have devised a schema for classifying forms of argument. In addition to emphasizing the logical relationships between parts of an argument, the schema adds a new dynamic to argument understanding theory by incorporating logical (substantive) and non-logical(authoritative, motivational) dimensions to the classificatory design. The Brockriede and Ehninger interpretation of the Toulmin model is useful for two main reasons in that it (a) provides a structure for analysis and criticism of rhetorical arguments and (b) suggests a typology for classifying arguments based on rhetorical proofs.

Toulmin's Construct

It is useful to elaborate the structure for argument analysis developed by Toulmin (1964). This model provides a two-tiered analysis of argument described as 'movement from accepted data through a warrant to a claim'. The first tier consists of data, warrant and claim.

Data - consist of facts that reflect events, statistical data, citations from authority etc. If data do not exist, an essential component of argument is missing, because there is no factual base and the argument fails to inform.

Claim - is the substance of the idea conveyed in an argument. It has been referred to as the main proof line by Brockriede & Enginger (1968). The claim can occur either as the final statement in an argument or it could be some midway position in an argument. The usual order is data first then claim. In this order, claim connotes 'therefore'. The reverse order infers 'because'.

Warrant - is the description of that component of the argument that moves the argument from data to claim. It certifies the proposition in the claim statement so that it becomes acceptable.

In addition to this first triad: data, warrant and claim, Toulmin's construct contains a second set of components of which all or some may exist in an argument. These are termed: backing, rebuttal and qualifier.

A backing is made up of measures or credentials designed to certify the beliefs of the warrant. This can be a single item or an entire argument within itself, complete with the first level triad of data, claim and warrant.

The rebuttal acts as a safety mechanism and is often attached to the claim statement. The rebuttal recognizes the cases in which the claim will be constrained, is invalid or is in some way in need of a qualifier. The rebuttal places limitations on the scope of the validity of a claim by foreseeing possible objections.

The role of the qualifier is to determine the strength of an argument. It does this by quantifying terms (e.g. 'probably') or, by pointing out possible fallacies. When an argument is thought to be axiomatic and cannot be disputed, no qualifier is attached.

Classification Approaches

Brockriede and Ehninger devised their classification system based on the attributes of rhetorical proof. The first type of argument, Substantive, is pictured as data which progress through to claim based on beliefs about things in the external world. The second, Authoritative, is grounded in structures about the quality of the source from which the data are derived. The third, Motivational, is based upon assumptions related to the emotive state, e.g. the ambitions, inner drives etc., of those who hear the argument. The assumptions of each class and its relationship to the two-level triad of elements for argument analysis are examined in turn.

1. Substantive Arguments - The substantive argument permits the discourse to move from from data to claim and is rooted in assumptions about the logical relationships that exist in the external world. There are six possible orderings of of this type of argument:

   a. Cause - In this type of argument, data comprise facts about an event, person , object or situation. The warrant provides the source of power for these facts and tells what effects they will have while claim relates these effects to the data.

   b. Sign - The data are made up of a set of symptoms. The warrant assigns appropriate meaning to these symptoms and claim explains the objects, people or situations that have these symptoms.

   c. Generalization - The data are made up of information about objects, persons, events or conditions representative of given items in a class of things. The warrant posits that this representative sample will extend to the whole population. The claim clarifies the underlying assumptions of the warrant.

   d. Parallel Case - The data are made up of one or more statements about a single object, event or condition. The warrant states that the case reported in the data resembles a second instance of the same category. The claim assumes that the new case can be likened to the first. A rebuttal in a parallel case argument

obtains when (a) another parallel case has a strong similarity to the existing case or (b) some strong dissimilarity disaffirms the warrant.

e. Analogy - The data report that a relationship of a certain nature exists between a pair of items. The claim amplifies the relationship implied in the warrant. The distinction between argument from parallel case and argument from analogy is that the former assumes resemblance between two cases while the latter assumes a similarity of relationship. Frequently, the relationship expressed in this type of argument requires the qualifier 'possibly'.

f. Classification - In this type of argument, the data reflect a generalized conclusion about known members of a class. The warrant assumes that the properties of the items under consideration can be extended to those items of the class which have yet to be examined. The claim gives to specific items the characteristics of the general statement. The provisos to argument from classification are (a) a class member may not share the special characteristics of the class specified in the data but may have enough other characteristics to justify membership in that class and (b) there may be instances when a given class member may not share the attributes of a class.

2. Authoritative Arguments - In this type of argument, data are composed of either factual reports or stated opinions. The warrant attests to the reliability of these reports or opinions. The claim reinforces the data statements whose credentials are supplied by the warrant.

3. Motivational Arguments - The data in this type of argument is composed of statements grounded in the claims of previous arguments. The warrant provides the rationale for accepting the claim by appealing to the some emotive quality, e.g. desire, in the hearer.

METHODOLOGY

In order to validate this classification scheme and the specific ordering of the different components of information in the various types of arguments defined above, a descriptive study was conducted on a sample of abstracts. Fifty-five argumentative type abstracts were selected from different issues of International Political Science Abstracts under the headings 'Political Thinkers and Ideas' and 'Methods and Theories'.

Each abstract was classified by the first author (Francis) into one of three major groupings: substantive, authoritative or motivational, based on the Brockriede and Ehninger (1968) classification. The choice was made by locating different kinds of propositions i.e. those of value, fact or of policy. The abstracts

reflecting each type of argument were subdivided into one of the six subcategories. Each abstract was examined for presence of both the first triad (data, warrant and claim) and second triad (backing, rebuttal and qualifier) of elements.

Linguistic Clues

The second stage of the investigation was an examination of lexical clues and verb tenses. In this context, clue words are those words and phrases which are used by writers to indicate explicitly how parts of the discourse should be interpreted. For our purposes, it was decided to concentrate on whether some of the clues to cohesion specified by Halliday and Hasan (1976) could be relied on to reveal which function of an argument was being performed by the different passages in each abstract.

Halliday and Hasan have outlined a taxonomy of types of cohesive relationships which conjoin one portion of text to another. One familiar type of explicitly marked cohesive relationship in texts is indicated by markers which relate what follows to what has been said before, e.g. 'and', 'but', 'so', 'then'. Halliday and Hasan (1976) provide an extended discussion of the relationships indicated by such markers. For example, the classes of explicit markers for conjunctive relations are:

- Additive - and, or, furthermore, similarly, in addition
- Adversative - but, however, on the other hand, nevertheless
- Causal - so, consequently, for this reason, it follows from this
- Temporal - then, after that, an hour later, finally, at last

A frequency analysis was performed on all words occurring in the 55 abstracts in an attempt to determine whether there was a skewed usage of conjunctive clues within different components of the argumentative abstracts.

RESULTS

Of the fifty-five abstracts, 38 (68%) were found to belong to the Substantive class, 11 (20%) to the Authoritative, while 6 (11%) were classed as Motivational. The Substantive class was subdivided into the six categories detailed above. Substantive/Sign arguments occurred most frequently (20/38 times) followed by Substantive/Cause 11/38 times. There were no abstracts classed as Substantive/Analogy.

Table 1 shows the number of occurrences of the two-level triad of elements with the three main classification groups as well as the sub-groups of the Substantive class. The first triad of elements (Data, Warrant, Claim) was found in all fifty-five abstracts. In the case of second triad, Rebuttal was found in 16 of the abstracts, while Qualifier was found in approximately half of the abstracts (24). Backing occurred 28 times.

Table 1: Frequency of 2-Level Triad of Elements

	1st Triad			2nd Triad		
	Data	Warrant	Claim	Back.	Rebutt.	Qualif.
Substantive (N = 38)						
Cause	11	11	11	7	-	5
Sign	20	20	20	9	11	11
Generalization	3	3	3	2	-	2
Parallel Case	3	3	3	3	1	2
Analogy	-	-	-	-	-	-
Classification	1	1	1	-	-	-
Authoritative (N = 11)	11	11	11	6	3	7
Motivational (N = 6)	6	6	6	1	1	3
	55	55	55	28	16	24

The results showed a reasonably predictable ordering of the structural components. Data preceded warrant (40/50) times; warrant preceded backing (33/55) times. An analysis was made of the observed orderings of the six components to detect what the most frequent order might be. As can be seen in Figure 1, there is no invariant ordering, although the most likely structure would be [DATA - WARRANT - BACKING - CLAIM]. There were no distinctive orderings for the three major classes of arguments established by Brockriede and Ehninger (1968). That is, substantive arguments, authoritative arguments, and motivational arguments do not differ significantly one from each other in terms of the ordering of argument components. The conjunctive relations suggested by Halliday and Hasan's work on cohesion were not found to be predictive of the particular function in an argument being played by the pieces of text.

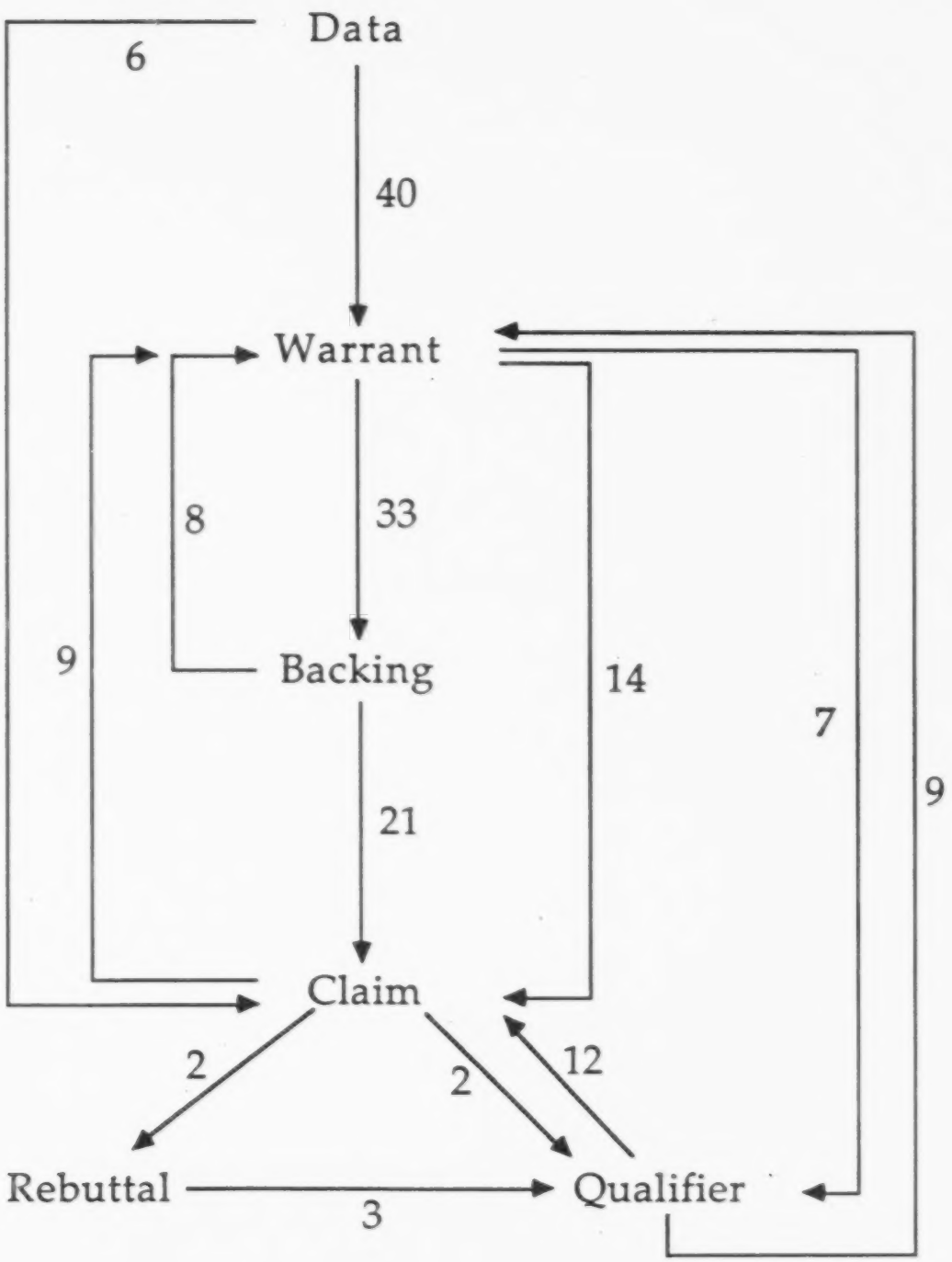

Figure 1: Observed Orderings of Components in 55 Theoretical Abstracts

CONCLUSIONS

The above presentation of the different types of argument definitely suggests that patterns of form and structure can be found in argumentation texts. These results strongly support the premise which inspired this investigation, namely that the rhetorical structure of argument is a useful framework in which to describe the structure of theoretical abstracts. In addition, since there were no significant differences between abstracts reflecting different types of arguments in terms of the orderings of components, it would appear that a single model will be suitable for describing the typical orderings of all such abstracts. However, the final goal of the researchers on the project, of which this study was a part, was to determine whether these orderings can be detected automatically by use of lexical clues. In this sense, the investigation was only partially successful. In order to present document representations which usefully identify the role that each portion of text performs in the argument, an automatic means of detecting these roles need to be identified. Although this research did not successfully establish the set of lexical clues which could be relied on to automatically structure theoretical abstracts, it did delineate the culturally validated discourse structure of argumentative discourse as being the structure underlying theoretical abstracts.

The importance of this discourse-level structure to information retrieval systems is that it offers the potential of a predictable structured representation in which the specific content of individual theoretical abstracts might be more easily evaluated by users. If an interface could take advantage of a structure like that in Figure 2, those users who are not very familiar with the topic on which they are seeking information would be aided in improving their comprehension of the retrieved documents. Perhaps even more importantly, these structured representations of abstracts might be useful to users as a means of refining their queries in the next iteration of the search.

AUTHORITATIVE

DATA According to Tocqueville, the most important determinant of the character of any society is the political culture (Moeurs).

CLAIM A political culture is shaped not only by sociological conditions and Laws, but also, in modern times, by ideas propounded by intellectuals.

WARRANT In Tocqueville's day, two dominant schools of thought were contending for influence over the public mind of Europe: philosopher rationalism and traditionalism

BACKING Neither one of these schools, Tocqueville argued, promoted a political culture that could reconcile liberty and democracy.

QUALIFIER Unlike the opposing schools, the new political science could not be propagated directly as an ideology.

BACKING Its implementation relied on an indexed strategy - using institutions to inculcate certain 'mental habits' among citizens. This in turn called for ways of limiting the role of intellectuals in influencing political culture.

Figure 1: Sample Abstract

REFERENCES

Anderson, J. and Dovre P. "The nature of rhetorical reasoning". IN Readings in Argumentation Boston: Allyn & Bacon, 1968.

Belkin, N.J., Oddy, R.N. & Brooks, H. "ASK for information retrieval: Part I. Background and theory. Part II. Results of a design study". Journal of Documentation, 38, 1982, 61-71, 145-64.

Brockriede, W. and Ehninger, D. "Toulmin on argument: an interpretation and Application". In Readings in Argumentation. Boston: Allyn & Bacon, 1968, 263-78.

Liddy, Elizabeth "The discourse level structure of natural language texts: an exploratory study of empirical abstracts". Diss. Syracuse University, 1988.

MacMullin, S. and Taylor, R. S. "Problem dimensions and information traits". The information society, 3, 1984, pp. 91-111.

McBurney, J. & Mills, G. "Probabilities". In Readings in Argumentation. Boston: Allyn & Bacon, 1968, pp. 240-42.

Taylor, R. S. "Question negotiation and information seeking in libraries". College and research libraries, 29, 1968, 178-94.

Toulmin S. The Uses of Argument. Cambridge: University Press, 1964, 264p.